Probab. Theory Relat. Fields (2007) 137:201–225
DOI 10.1007/s00440-006-0026-1

Concentration inequalities for random fields via coupling

J.-R. Chazottes · P. Collet · C. Külske · F. Redig

Received: 28 March 2005 / Revised: 28 June 2006 /
Published online: 22 September 2006

Abstract We present a new and simple approach to concentration inequalities in the context of dependent random processes and random fields. Our method is based on coupling and does not use information inequalities. In case one has a uniform control on the coupling, one obtains exponential concentration inequalities. If such a uniform control is no more possible, then one obtains polynomial or stretched-exponential concentration inequalities. Our abstract results apply to Gibbs random fields, both at high and low temperatures and in particular to the low-temperature Ising model which is a concrete example of non-uniformity of the coupling.

Keywords Exponential concentration · Stretched-exponential concentration · Moment inequality · Gibbs random fields · Ising model · Orlicz space · Luxembourg norm · Kantorovich–Rubinstein theorem

J.-R. Chazottes · P. Collet
Centre de Physique Théorique, CNRS UMR 7644,
91128 Palaiseau Cedex, France
e-mail: jeanrene@cpht.polytechnique.fr

P. Collet
e-mail: collet@cpht.polytechnique.fr

C. Külske
Department of Mathematics and Computing Sciences, University of Groningen,
Blauwborgje 3, 9747 AC Groningen, The Netherlands
e-mail: kuelske@math.rug.nl

F. Redig (✉)
Mathematisch Instituut Universiteit Leiden, Niels Bohrweg 1,
2333 CA Leiden, The Netherlands
e-mail: redig@math.leidenuniv.nl

Mathematical Subject Classification (2000) 60k35 (Primary); 82c22 (Secondary)

1 Introduction

By now, concentration inequalities for product measures have become a standard and powerful tool in many areas of probability and statistics, such as density estimation [5], geometric probability [24], etc. A recent monograph about this area is [12] where the reader can find much more information and relevant references. Exponential concentration inequalities for functions of dependent, strongly mixing random variables were obtained for instance in [11,15–17,20,21]. In the context of dynamical systems, Collet et al. [3] obtained an exponential concentration inequality for separately Lipschitz functions using spectral analysis of the transfer operator. Külske [11] obtained an exponential concentration inequality for functions of Gibbs random fields in the Dobrushin uniqueness regime. Therein the main input is Theorem 8.20 in [9] which allows to estimate uniformly the terms appearing in the martingale difference decomposition in terms of the Dobrushin matrix. Marton [16] obtained exponential concentration results for a class of Gibbs random fields under a strong mixing condition lying between Dobrushin–Shlosman condition and its weakening in the sense of E. Olivieri, P. Picco and F. Martinelli.

Besides exponential concentration inequalities, polynomial concentration inequalities easily follow from upper bounds on moments. In the context of product measures, bounds on the variance are well-known [4,5]. In the context of dynamical systems, a bound on the variance is obtained in [2].

The approach followed in [15–17,21] uses coupling ideas and information inequalities, such as Pinsker inequality. Such inequalities can only lead to exponential concentration inequalities. This can be understood easily since it is well-known [1] that there is equivalence between information inequalities and exponential inequalities on the Laplace transform, the latter yielding exponential concentration inequalities by Chebychev's inequality.

The purpose of the present paper is to derive abstract bounds allowing to obtain not only exponential, but also polynomial and stretched-exponential concentration inequalities. In particular, this means that we do not use information inequalities. Going beyond the exponential case was motivated by the low-temperature Ising model which can not satisfy an exponential concentration inequality for the magnetization. Here we obtain abstract concentration inequalities using a coupling approach. Our setting is (dependent) random variables indexed by $\mathbb{Z}^d$, $d \geq 1$, and taking values in a finite alphabet. We are interested in obtaining concentration inequalities for "local" functions g around their expectation $\mathbb{E}g$ in terms of their variations. The inter-dependence between random variables is measured by a "coupling matrix" which tells us how "well" one can couple in the far "future" if the "past" is given. If the coupling matrix can be uniformly controlled in the realization, then an exponential concentration inequality follows. If the coupling matrix cannot be controlled uniformly in the realization, then we typically obtain bounds for moments and for Luxembourg norms of $g - \mathbb{E}g$. In the former case this leads to polynomial concentration inequalities, in the latter case this gives stretched-exponential concentration inequalities.

As a first application of our abstract inequalities, we obtain an exponential concentration inequality for Gibbs random fields in a "high-temperature" regime, complementary to the Dobrushin uniqueness regime studied in [11]. A second application is the "low-temperature" Ising model for which the coupling matrix cannot be uniformly controlled in the realization, and for which the previous methods [16,21] do not apply. We obtain polynomial, even stretched-exponential, concentration inequalities for the low-temperature Ising model. Let us mention that our concentration inequalities yield various non-trivial applications which will be the subject of a forthcoming paper.

The paper is organized as follows. In Sect. 2, we state and prove our abstract inequalities, first in the context of random fields indexed by $\mathbb{Z}$, and next when the index set is $\mathbb{Z}^d$, $d \geq 2$. Section 3 deals with high-temperature Gibbs measures and the low-temperature Ising model.

2 Main results

Let A be a finite set. Let $g : A^n \to \mathbb{R}$ be a function of n-variables. An element σ of the set $A^{\mathbb{N}}$ is an infinite sequence drawn from A, i.e., $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_i, \dots)$ where $\sigma_i \in A$. With a slight abuse of notation, we also consider g as a function on $A^{\mathbb{N}}$ which does not depend on σ_k, for all $k > n$.

A concentration inequality is an estimate for the probability of concentration of the function g from its expectation, i.e., an estimate for

$$\mathbb{P}\{|g - \mathbb{E}g| \geq t\} \tag{1}$$

for all $n \geq 1$ and all $t > 0$, within a certain class of probability measures $\mathbb{P}$. For example, an *exponential concentration inequality* is obtained by estimating the expectation

$$\mathbb{E}[e^{\lambda(g - \mathbb{E}g)}]$$

for any $\lambda \in \mathbb{R}$, and using the exponential Chebychev's inequality.

However, there are natural examples where the exponential concentration inequality does not hold (see the example of the low-temperature Ising model below). In that case we are interested in bounding moments of the form

$$\mathbb{E}[(g - \mathbb{E}g)^{2p}]$$

to control the probability (1).

In this section, we use a combination of the classical martingale decomposition of $g - \mathbb{E}g$ and maximal coupling to perform a further telescoping which is adequate for the dependent case. This will lead us to a "coupling matrix" depending on the realization $\sigma \in A^{\mathbb{N}}$. This matrix quantifies how "good" future symbols can be coupled if past symbols are given according to σ. Typically, we have in mind applications to Gibbs random fields. In that framework, the elements of the coupling matrix can be controlled uniformly in σ in the "high-temperature regime". This uniform control leads naturally to an exponential concentration inequality. At low temperature we can only control the coupling matrix for "good" configurations, but not uniformly. Therefore an exponential concentration inequality cannot hold (for all g). Instead we will obtain polynomial and stretched-exponential concentration inequalities. This will be done by controlling moments and Luxembourg norms of $g - \mathbb{E}g$.

2.1 The coupling matrix D^σ

We now present our method. For $i = 1, 2, \ldots, n$, let $\mathcal{F}_i$ be the sigma-field generated by the random variables $\sigma_1, \ldots, \sigma_i$, and $\mathcal{F}_0$ be the trivial sigma-field $\{\emptyset, \Omega\}$. We write

$$g(\sigma_1, \ldots, \sigma_n) - \mathbb{E}g = \sum_{i=1}^{n} V_i(\sigma), \tag{2}$$

where

$$\begin{aligned}
V_i(\sigma) &:= \mathbb{E}[g|\mathcal{F}_i](\sigma) - \mathbb{E}[g|\mathcal{F}_{i-1}](\sigma) \\
&= \int \mathbb{P}(\mathrm{d}\eta_{i+1} \cdots \mathrm{d}\eta_n | \sigma_1, \ldots, \sigma_i)\, g(\sigma_1, \ldots, \sigma_i, \eta_{i+1}, \ldots, \eta_n) \\
&\quad - \int \mathbb{P}(\mathrm{d}\eta_i \cdots \mathrm{d}\eta_n | \sigma_1, \ldots, \sigma_{i-1})\, g(\sigma_1, \ldots, \sigma_{i-1}, \eta_i, \eta_{i+1}, \ldots, \eta_n) \\
&= \int \mathbb{P}(\mathrm{d}\eta_{i+1} \cdots \mathrm{d}\eta_n | \sigma_1, \ldots, \sigma_i)\, g(\sigma_1, \ldots, \sigma_i, \eta_{i+1}, \ldots, \eta_n) \\
&\quad - \int \mathbb{P}(\mathrm{d}\eta_i | \sigma_1, \ldots, \sigma_{i-1}) \int \mathbb{P}(\mathrm{d}\eta_{i+1} \cdots \mathrm{d}\eta_n | \sigma_1, \ldots, \sigma_{i-1}, \eta_i)\, g \\
&\quad \times (\sigma_1, \ldots, \sigma_{i-1}, \eta_i, \eta_{i+1}, \ldots, \eta_n) \\
&\leq \max_{a \in A} \int \mathbb{P}(\mathrm{d}\eta_{i+1} \cdots \mathrm{d}\eta_n | \sigma_1, \ldots, \sigma_{i-1}, \sigma_i = a) \\
&\quad \times g(\sigma_1, \ldots, \sigma_{i-1}, a, \eta_{i+1}, \ldots, \eta_n) \\
&\quad - \min_{b \in A} \int \mathbb{P}(\mathrm{d}\eta_{i+1} \cdots \mathrm{d}\eta_n | \sigma_1, \ldots, \sigma_{i-1}, \sigma_i = b) \\
&\quad \times g(\sigma_1, \ldots, \sigma_{i-1}, b, \eta_{i+1}, \ldots, \eta_n) \\
&=: Y_i(\sigma) - X_i(\sigma).
\end{aligned} \tag{3}$$

Denote by $\mathbb{P}^\sigma_{i,a,b} = \mathbb{P}^{\sigma_{<i}}_{i,a,b}$ the maximal coupling [13] of the conditional distributions $\mathbb{P}(d\eta_{\geq i+1} | \sigma_1, \ldots, \sigma_{i-1}, \sigma_i = a)$ and $\mathbb{P}(d\eta_{\geq i+1} | \sigma_1, \ldots, \sigma_{i-1}, \sigma_i = b)$, that is, the coupling $\mathbb{Q}$ for which the expected distance

$$\int d\left(\eta^{(1)}_{j\geq i+1}, \eta^{(2)}_{j\geq i+1}\right) d\mathbb{Q}(\eta^{(1)}, \eta^{(2)})$$

is minimal and equal to the Vasserstein distance between $\mathbb{P}(d\eta_{\geq i+1}|\sigma_1, \ldots, \sigma_{i-1}, \sigma_i = a)$ and $\mathbb{P}(d\eta_{\geq i+1}|\sigma_1, \ldots, \sigma_{i-1}, \sigma_i = b)$ (see also [6]). Now we introduce the (infinite) upper-triangular matrix D^σ defined for $i, j \in \mathbb{N}$ by

$$\begin{aligned} D^\sigma_{ii} &:= 1 \\ D^\sigma_{i,i+j} &:= \max_{a,b\in A} \mathbb{P}^\sigma_{i,a,b}\left\{\sigma^{(1)}_{i+j} \neq \sigma^{(2)}_{i+j}\right\}. \end{aligned} \tag{4}$$

Notice that if the σ_i's are mutually independent, then D^σ is the identity matrix because the conditional distributions $\mathbb{P}(d\eta_{\geq i+1}|\sigma_1, \ldots, \sigma_{i-1}, \sigma_i = a)$ and $\mathbb{P}(d\eta_{\geq i+1}|\sigma_1, \ldots, \sigma_{i-1}, \sigma_i = b)$ are equal. Hence we have a perfect coupling in this case.

The matrix D^σ is the analogue of the matrix γ introduced in [21], but here we keep the σ-dependence and do not take straight away the supremum over σ. This has the advantage that "exceptional" σ for which the matrix elements $D^\sigma_{i,i+j}$ do not decay as j grows, can be dealt with by averaging "at the end."

We proceed with the following simple telescoping identity:

$$\begin{aligned} & g(\sigma_1, \ldots, \sigma_{i-1}, a, \sigma^{(1)}_{i+1}, \ldots, \sigma^{(1)}_n) - g(\sigma_1, \ldots, \sigma_{i-1}, b, \sigma^{(2)}_{i+1}, \ldots, \sigma^{(2)}_n) \\ &\quad = [g(\sigma_1, \ldots, \sigma_{i-1}, a, \sigma^{(1)}_{i+1}, \ldots, \sigma^{(1)}_n) - g(\sigma_1, \ldots, \sigma_{i-1}, b, \sigma^{(1)}_{i+1}, \ldots, \sigma^{(1)}_n)] \\ &\qquad + [g(\sigma_1, \ldots, \sigma_{i-1}, b, \sigma^{(1)}_{i+1}, \ldots, \sigma^{(1)}_n) - g(\sigma_1, \ldots, \sigma_{i-1}, b, \sigma^{(2)}_{i+1}, \sigma^{(1)}_{i+2}, \ldots, \sigma^{(1)}_n)] \\ &\qquad + [g(\sigma_1, \ldots, \sigma_{i-1}, b, \sigma^{(2)}_{i+1}, \sigma^{(1)}_{i+2}, \ldots, \sigma^{(1)}_n) - g \\ &\qquad \times (\sigma_1, \ldots, \sigma_{i-1}, b, \sigma^{(2)}_{i+1}, \sigma^{(2)}_{i+2}, \sigma^{(1)}_{i+3}, \ldots, \sigma^{(1)}_n)] \\ &\qquad + \cdots + [g(\sigma_1, \ldots, \sigma_{i-1}, b, \sigma^{(2)}_{i+1}, \sigma^{(2)}_{i+2}, \ldots, \sigma^{(2)}_{n-1}, \sigma^{(1)}_n) - g \\ &\qquad \times (\sigma_1, \ldots, \sigma_{i-1}, b, \sigma^{(2)}_{i+1}, \ldots, \sigma^{(2)}_n)] \\ &\quad =: \sum_{j=0}^{n-i} \nabla^{12}_{i,i+j} g. \end{aligned}$$

We define the variation of g at site i by

$$\delta_i g := \sup_{\substack{\sigma_j = \sigma'_j \\ \forall j \neq i}} |g(\sigma) - g(\sigma')|,$$

and by construction we have the inequality

$$\nabla^{12}_{i,i+j} g \leq \delta_{i+j} g \; \mathbb{1}_{\sigma^{(1)}_{i+j} \neq \sigma^{(2)}_{i+j}}.$$

It follows from (3) and (4) that

$$\begin{aligned} Y_i(\sigma) - X_i(\sigma) &= \max_{a,b\in A} \left\{ \int \mathbb{P}(\mathrm{d}\eta_{i+1}\cdots \mathrm{d}\eta_n | \sigma_1,\ldots,\sigma_{i-1},\sigma_i = a) \right. \\ &\quad \times g(\sigma_1,\ldots,\sigma_{i-1},a,\eta_{i+1},\ldots,\eta_n) \\ &\quad - \int \mathbb{P}(\mathrm{d}\eta_{i+1}\cdots \mathrm{d}\eta_n | \sigma_1,\ldots,\sigma_{i-1},\sigma_i = b) \\ &\quad \left. \times g(\sigma_1,\ldots,\sigma_{i-1},b,\eta_{i+1},\ldots,\eta_n) \right\} \\ &= \max_{a,b\in A} \left\{ \int \mathbb{P}^{\sigma}_{i,a,b}(\mathrm{d}\sigma^{(1)}_{\geq i+1}, \mathrm{d}\sigma^{(2)}_{\geq i+1}) \big[g(\sigma_1,\ldots,\sigma_{i-1},a,\sigma^{(1)}_{i+1},\ldots,\sigma^{(1)}_n) \right. \\ &\quad \left. - g(\sigma_1,\ldots,\sigma_{i-1},b,\sigma^{(2)}_{i+1},\ldots,\sigma^{(2)}_n) \big] \right\} \\ &\leq \max_{a,b\in A} \sum_{j=0}^{n-i} \delta_{i+j} g \, \mathbb{P}^{\sigma}_{i,a,b} \left\{ \sigma^{(1)}_{i+j} \neq \sigma^{(2)}_{i+j} \right\} \\ &\leq \sum_{j=0}^{n-i} D^{\sigma}_{i,i+j} \, \delta_{i,i+j} g = (D^{\sigma}\delta g)_i, \end{aligned}$$

where δg denotes the column vector with coordinates $\delta_j g$, for $j = 1,\ldots,n$, and 0 for $j > n$. Therefore, we get the inequality

$$V_i(\sigma) = Y_i(\sigma) - X_i(\sigma) \leq (D^{\sigma}\delta g)_i \,. \tag{5}$$

Applying the above reasoning to $-g$ shows that the previous inequality also applies to $-V_i$.

Remark 1 The advantage of the previous bound is that it only involves δg. One could imagine to consider, for instance, the second moment of $\nabla^{12}_{i,i+j} g$ instead. This could lead to better results but it has the drawback that we need to know much more about the coupling than we usually do.

2.2 Uniform decay of D^{σ}: exponential concentration inequality

Let $\overline{D}_{i,j} := \sup_{\sigma\in A^{\mathbb{N}}} D^{\sigma}_{i,j}$.

We assume that the following operator $\ell^2(\mathbb{N})$-norm is finite:

$$\|\overline{D}\|^2_{\ell^2(\mathbb{N})} := \sup_{u\in\ell_2(\mathbb{N}), \|u\|_{\ell^2(\mathbb{N})}=1} \|\overline{D}u\|^2_{\ell^2(\mathbb{N})} < \infty \,. \tag{6}$$

We have the following exponential concentration inequality.

Theorem 1 *Let $n \in \mathbb{N}$ be arbitrary. Assume that* (6) *holds. Then, for all functions $g : A^n \to \mathbb{R}$, we have the inequality*

$$\mathbb{P}\{|g - \mathbb{E}g| \geq t\} \leq 2\exp\left(-\frac{2t^2}{\|\overline{D}\|^2_{\ell^2(\mathbb{N})}\ \|\delta g\|^2_{\ell^2(\mathbb{N})}}\right). \tag{7}$$

for all $t > 0$.

Proof We recall the following lemma which is proved in [5].

Lemma 1 *Suppose $\mathcal{F}$ is a sigma-field and Z_1, Z_2, V are random variables such that*

1. $Z_1 \leq V \leq Z_2$
2. $\mathbb{E}(V|\mathcal{F}) = 0$
3. *Z_1 and Z_2 are $\mathcal{F}$-measurable.*

Then, for all $\lambda \in \mathbb{R}$, we have

$$\mathbb{E}(\mathrm{e}^{\lambda V}|\mathcal{F}) \leq \mathrm{e}^{\lambda^2(Z_2 - Z_1)^2/8}. \tag{8}$$

We apply this lemma with $V = V_i$, $\mathcal{F} = \mathcal{F}_{i-1}$,
$Z_1 = X_i - \mathbb{E}[g|\mathcal{F}_{i-1}]$, $Z_2 = Y_i - \mathbb{E}[g|\mathcal{F}_{i-1}]$. *Using inequality* (5)

$$V_i(\sigma) = Y_i(\sigma) - X_i(\sigma) \leq (D^\sigma \delta g)_i,$$

we obtain

$$\mathbb{E}(\mathrm{e}^{\lambda V_i}|\mathcal{F}_{i-1})(\sigma) \leq \mathrm{e}^{\lambda^2 (D^\sigma \delta g)_i^2/8}. \tag{9}$$

Therefore, by successive conditioning, and the exponential Chebychev's inequality,

$$\begin{aligned}
\mathbb{P}\{g - \mathbb{E}g \geq t\} &\leq \mathrm{e}^{-\lambda t}\mathbb{E}\left(\mathrm{e}^{\lambda\sum_{i=1}^n V_i}\right)\\
&\leq \mathrm{e}^{-\lambda t}\mathbb{E}\left(\mathbb{E}(e^{\lambda V_n}|\mathcal{F}_{n-1}) e^{\lambda \sum_{i=1}^{n-1} V_i}\right)\\
&\leq \cdots \leq \mathrm{e}^{-\lambda t}\exp\left(\frac{\lambda^2}{8}\|\overline{D}\delta g\|^2_{\ell^2(\mathbb{N})}\right)\\
&\leq \mathrm{e}^{-\lambda t}\exp\left(\frac{\lambda^2}{8}\|\overline{D}\|^2_{\ell^2(\mathbb{N})}\ \|\delta g\|^2_{\ell^2(\mathbb{N})}\right).
\end{aligned} \tag{10}$$

Now choose the optimal $\lambda = 4t/(\|\overline{D}\|^2_{\ell^2(\mathbb{N})}\ \|\delta g\|_2^2)$ to obtain

$$\mathbb{P}\{g - \mathbb{E}g \geq t\} \leq \exp\left(-\frac{2t^2}{\|\overline{D}\|^2_{\ell^2(\mathbb{N})}\ \|\delta g\|^2_{\ell^2(\mathbb{N})}}\right).$$

Combining the inequality for g and the one for −g yields (7). *The theorem is proved.*

2.3 Non-uniform decay of D^σ: polynomial and stretched-exponential concentration inequalities

If the dependence on σ of the elements of the coupling matrix cannot be controlled uniformly, then in many cases we can still control the moments of the coupling matrix. To this aim, we introduce the (non-random, i.e., not depending on σ) matrices

$$\mathcal{D}^{(p)}_{i,j} := \mathbb{E}[(D^\sigma_{i,j})^p]^{1/p} \tag{11}$$

for all $p \in \mathbb{N}$.

A typical example of non-uniformity which we will encounter, for instance in the low-temperature Ising model, is an estimate of the following form:

$$D^\sigma_{i,i+j} \leq 1\!\!1\{\ell_i(\sigma) \geq j\} + \psi_j, \tag{12}$$

where $\psi_j \geq 0$ does not depend on σ, and where ℓ_i are unbounded functions of σ with a distribution independent of i. The idea is that the matrix elements $D^\sigma_{i,i+j}$ "start to decay" when $j \geq \ell_i(\sigma)$. The "good" configurations σ are those for which $\ell_i(\sigma)$ is "small".

In the particular case when (12) holds, in principle one still can have an exponential concentration inequality provided one is able to bound

$$\mathbb{E}\left(\mathrm{e}^{\lambda\sum_{i=1}^n \ell_i^2}\right).$$

However, in the example given below, the tail of the ℓ_i will be stretched exponential. Henceforth, we cannot deduce an exponential concentration inequality from these estimates.

We now prove an inequality for the variance of g which is a generalization of an inequality derived in [4] in the i.i.d. case.

Theorem 2 *Let $n \in \mathbb{N}$ be arbitrary. Then for all functions $g : A^n \to \mathbb{R}$ we have the inequality*

$$\mathbb{E}\left[(g - \mathbb{E}g)^2\right] \leq \|\mathcal{D}^{(2)}\|^2_{\ell^2(\mathbb{N})}\,\|\delta g\|^2_{\ell^2(\mathbb{N})}. \tag{13}$$

Proof We start again from the decomposition (2). Recall the fact that $\mathbb{E}[V_i|\mathcal{F}_j] = 0$ for all $i > j$, from which it follows that $\mathbb{E}[V_iV_j] = 0$ for $i \neq j$.

Using (5) and Cauchy-Schwarz's inequality we obtain

$$\begin{aligned}
\mathbb{E}\left[(g-\mathbb{E}g)^2\right] &= \mathbb{E}\sum_{i=1}^{n} V_i^2 \\
&\le \mathbb{E}\left(\sum_{i=1}^{n}(D\delta g)_i^2\right) \\
&= \sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathbb{E}\left(D_{i,k}D_{i,l}\right)\delta_k g\delta_l g \\
&\le \sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathbb{E}\left(D_{i,k}^2\right)^{\frac{1}{2}}\mathbb{E}\left(D_{i,l}^2\right)^{\frac{1}{2}}\delta_k g\delta_l g \\
&= \|\mathcal{D}^{(2)}\delta g\|_{\ell^2(\mathbb{N})}^2 \\
&\le \|\mathcal{D}^{(2)}\|_{\ell^2(\mathbb{N})}^2\ \|\delta g\|_{\ell^2(\mathbb{N})}^2 .
\end{aligned}$$

Remark 2 In the i.i.d. case, the coupling matrix D is the identity matrix. Hence inequality (13) reduces to

$$\mathbb{E}\left[(g-\mathbb{E}g)^2\right] \le \|\delta g\|_{\ell^2(\mathbb{N})}^2$$

which is the analogue of Theorem 4 in [4].

We now turn to higher moment estimates. We have the following theorem from which we recover Theorem 2 but with a bigger constant.

Theorem 3 *Let $n \in \mathbb{N}$ be arbitrary. For all functions $g : A^n \to \mathbb{R}$ and for any $p \in \mathbb{N}$, we have*

$$\mathbb{E}\left[(g-\mathbb{E}g)^{2p}\right] \le (20p)^{2p}\ \|\mathcal{D}^{(2p)}\|_{\ell^2(\mathbb{N})}^{2p}\|\delta g\|_{\ell^2(\mathbb{N})}^{2p}.$$

Proof We start from (2) and get

$$\mathbb{E}\left[(g-\mathbb{E}g)^{2p}\right] = \mathbb{E}\left[\left(\sum_{i=1}^{n} V_i\right)^{2p}\right].$$

Now, by (2) and since $\mathbb{E}(V_i|\mathcal{F}_j) = 0$ for $i > j$, $M_k := \sum_{i=1}^{k} V_i$ defines a martingale w.r.t. the filtration $\mathcal{F}_k$, with $M_n = g - \mathbb{E}g$. Therefore, application of the Burkholder-Gundy's inequality [7, formula II.2.8, p. 41], gives for all $q \ge 2$

$$\mathbb{E}\left[|g-\mathbb{E}g|^q\right]^{\frac{1}{q}} \le 10q\ \mathbb{E}\left[\left(\sum_{i=1}^{n} V_i^2\right)^{\frac{q}{2}}\right]^{\frac{1}{q}}.$$

Therefore, for $q = 2p$, $p \in \mathbb{N}$, this gives at once

$$\mathbb{E}\left[(g - \mathbb{E}g)^{2p}\right] \leq (20p)^{2p}\, \mathbb{E}\left[\left(\sum_{i=1}^{n} V_i^2\right)^p\right].$$

We now estimate the rhs by using (5):

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_i V_i^2\right)^p\right] &= \sum_{i_1}\cdots\sum_{i_p}\mathbb{E}\left(V_{i_1}^2\cdots V_{i_p}^2\right) && (14)\\
&\leq \sum_{i_1}\cdots\sum_{i_p}\mathbb{E}\left[(D\delta g)_{i_1}^2\cdots(D\delta g)_{i_p}^2\right]\\
&= \sum_{i_1\cdots i_p}\sum_{j_1\cdots j_p}\sum_{k_1\cdots k_p}\mathbb{E}\left(\prod_{r=1}^{p} D_{i_r,j_r}D_{i_r,k_r}\right)\left(\prod_{r=1}^{p}\delta_{j_r}g\,\delta_{k_r}g\right)\\
&\leq \sum_{i_1\cdots i_p}\sum_{j_1\cdots j_p}\sum_{k_1\cdots k_p}\prod_{r=1}^{p}\left(\mathcal{D}^{(2p)}_{i_r,j_r}\mathcal{D}^{(2p)}_{i_r,k_r}\delta_{j_r}g\,\delta_{k_r}g\right)\\
&= \|\mathcal{D}^{(2p)}\delta g\|^{2p}_{\ell^2(\mathbb{N})} \leq \|\mathcal{D}^{(2p)}\|^{2p}_{\ell^2(\mathbb{N})}\,\|\delta g\|^{2p}_{\ell^2(\mathbb{N})}, && (15)
\end{aligned}$$

where in the fourth step we used the inequality

$$\mathbb{E}(f_1\cdots f_{2p}) \leq \prod_{i=1}^{2p}(\mathbb{E}(f_i^{2p}))^{\frac{1}{2p}},$$

which follows from Hölder's inequality. □

In order to be able to apply Theorems 2 and 3, one needs to estimate $\|\mathcal{D}^{(2p)}\|_{\ell^2(\mathbb{N})}$.

Proposition 1 *Assume inequality* (12) *holds, and let* $p \in \mathbb{N}$. *We have the bound*

$$\|\mathcal{D}^{(2p)}\|_{\ell^2(\mathbb{N})} \leq \sum_{j=1}^{\infty}\mathbb{P}\big(\ell_0(\sigma) \geq j\big)^{1/2p} + \|\psi\|_{\ell^1(\mathbb{N})}.$$

Proof We start by an upper estimate of $\mathcal{D}^{(2p)}$. From the definition (11) and the bound (12) we have using Minkowski's inequality (for $j \geq i$)

$$\begin{aligned}
\mathcal{D}^{(2p)}_{i,j} &= \mathbb{E}\big[(D^{\sigma}_{i,j})^{2p}\big]^{1/2p} \leq \mathbb{E}\Big(\big(\mathbb{1}\{\ell_i(\sigma) \geq j-i\} + \psi_{j-i}\big)^{2p}\Big)^{\frac{1}{2p}}\\
&\leq \mathbb{E}\left[\big(\mathbb{1}\{\ell_i(\sigma) \geq j-i\}\big)^{2p}\right]^{\frac{1}{2p}} + \psi_{j-i} \leq \mathbb{P}\big(\ell_0(\sigma) \geq j-i\big)^{\frac{1}{2p}} + \psi_{j-i} =: u_{i-j}\,, && (16)
\end{aligned}$$

since the law of ℓ_i is independent of i.

Now take $v \in \ell^2(\mathbb{N})$ with $\|v\|_{\ell^2(\mathbb{N})} = 1$. We have

$$\|\mathcal{D}^{(2p)} v\|_{\ell^2(\mathbb{N})} \leq \left\| \sum_{k=1}^{\infty} \mathcal{D}_{i,k}^{(2p)} |v_k| \right\|_{\ell^2(\mathbb{N})} \leq \left\| \sum_{k=1}^{\infty} u(i-k)|v_k| \right\|_{\ell^2(\mathbb{N})},$$

where the second inequality comes from (16). Since we have the $\ell^2(\mathbb{N})$-norm of a convolution, we can apply Young's inequality (see, e.g., [25]) to get

$$\|\mathcal{D}^{(2p)}\|_{\ell^2(\mathbb{N})} \leq \|u\|_{\ell^1(\mathbb{N})}.$$

The result immediately follows. □

Before we state the next theorem, which is a corollary of Proposition 1 and Theorem 3, we need the definition of some Orlicz spaces. We only deal here with a restricted class useful in our applications, we refer to [19,25] for the general definition. For $\varrho > 0$, let $\Phi_\varrho : \mathbb{R} \to \mathbb{R}^+$ be the Young function defined by

$$\Phi_\varrho(x) = \mathrm{e}^{(|x|+h_\varrho)^\varrho} - \mathrm{e}^{h_\varrho^\varrho},$$

where $h_\varrho = ((1-\varrho)/\varrho)^{1/\varrho} \mathbb{1}\{0 < \varrho < 1\}$. These are the Young functions used in particular in [22]. We recall that (see [25]) the Luxembourg norm with respect to Φ_ϱ of a random variable Z is defined by

$$\|Z\|_{\Phi_\varrho} = \inf\left\{\lambda > 0 \,\middle|\, \mathbb{E}\left(\Phi_\varrho\left(\frac{Z}{\lambda}\right)\right) \leq 1\right\}.$$

Remark 3 Note that for $\Phi_p(x) = |x|^p$, the Luxembourg norm is nothing but the L^p-norm.

Theorem 4 *Let $n \in \mathbb{N}$ be arbitrary. Then, for all functions $g : A^n \to \mathbb{R}$, for any $p \in \mathbb{N}$ and any $\epsilon > 0$, we have*

$$\begin{aligned}&\mathbb{E}\big[(g - \mathbb{E}g)^{2p}\big] \\ &\quad\leq (20p)^{2p} \left(\zeta(1+\epsilon/(2p-1))^{(2p-1)/2p} \mathbb{E}\big(\ell_0^{2p+\epsilon}\big)^{1/2p} + \|\psi\|_{\ell^1(\mathbb{N})}\right)^{2p} \|\delta g\|_{\ell^2(\mathbb{N})}^{2p}\end{aligned} \tag{17}$$

where ζ denotes Riemann's zeta function. For any $\vartheta > 0$, there is a constant $C_\vartheta > 0$, such that for any $\varrho < \vartheta/(1+\vartheta)$ satisfying $\zeta(\vartheta(1-\varrho)/\varrho) \geq 1$, we have

$$\left\| \frac{g - \mathbb{E}g}{\|\delta g\|_{\ell^2(\mathbb{N})}} \right\|_{\Phi_\varrho} \leq C_\vartheta \left(\zeta\left(\frac{\vartheta(1-\varrho)}{\varrho}\right) \|\ell_0\|_{\Phi_\vartheta}^{\vartheta(1-\varrho)/\varrho} + \|\psi\|_{\ell^1(\mathbb{N})} \right). \tag{18}$$

Remark 4 A similar result holds when $\zeta(\vartheta(1-\varrho)/\varrho) < 1$ with the square root of the zeta function. Note also that when ϱ increases to $\vartheta/(1+\vartheta)$, the number $\vartheta(1-\varrho)/\varrho$ decreases to one.

Proof We first estimate $\|\mathcal{D}^{(2p)}\|_{\ell^2(\mathbb{N})}$ in terms of some moment of ℓ_0. Let $\epsilon' = \epsilon/(2p-1)$. We have using Hölder inequality

$$\begin{aligned}\sum_{j=1}^{\infty} \mathbb{P}\big(\ell_0(\sigma) \geq j\big)^{1/2p} &= \sum_{j=1}^{\infty} j^{(2p-1)(1+\epsilon')/2p} \mathbb{P}\big(\ell_0(\sigma) \geq j\big)^{1/2p}\, j^{-(2p-1)(1+\epsilon')/2p} \\ &\leq \zeta(1+\epsilon')^{(2p-1)/2p} \left(\sum_{j=1}^{\infty} j^{2p-1+\epsilon} \mathbb{P}\big(\ell_0(\sigma) \geq j\big)\right)^{1/2p} \\ &\leq \zeta(1+\epsilon')^{(2p-1)/2p}\, \mathbb{E}\big(\ell_0^{2p+\epsilon}\big)^{1/2p} .\end{aligned}$$

Using Proposition 1 we get

$$\|\mathcal{D}^{(2p)}\|_{\ell^2(\mathbb{N})} \leq \zeta(1+\epsilon')^{(2p-1)/2p}\, \mathbb{E}\big(\ell_0^{2p+\epsilon}\big)^{1/2p} + \|\psi\|_{\ell^1(\mathbb{N})},$$

and (17) follows using Theorem 3.

To prove (18), we first observe that from (17) we have, for q even

$$\left\| \frac{g-\mathbb{E}g}{\|\delta g\|_{\ell^2(\mathbb{N})}} \right\|_{L^q(\mathbb{P})} \leq 10q\Big(\zeta(1+\epsilon/(q-1))^{\frac{q-1}{q}}\, \mathbb{E}(\ell_0^{q+\epsilon})^{\frac{1}{q}} + \|\psi\|_{\ell^1(\mathbb{N})}\Big).$$

We now recall that for any $1 > \varrho > 0$, there is a constant $\tilde{C}_\varrho > 1$ such that

$$\tilde{C}_\varrho^{-1} \sup_{q>2} \frac{\|Z\|_{L^q(\mathbb{P})}}{q^{1/\varrho}} \leq \|Z\|_{\Phi_\varrho} \leq \tilde{C}_\varrho \sup_{q>2} \frac{\|Z\|_{L^q(\mathbb{P})}}{q^{1/\varrho}}.$$

(See, e.g., [22] for a proof.) It is easy to verify using Young's inequality that the same inequality holds (with slightly different constants) when the supremum is taken over the q even integers, and we will only consider such q below.

Therefore if $0 < \varrho < \vartheta/(1+\vartheta)$, taking

$$\epsilon = \vartheta q\left(\frac{1}{\varrho} - \frac{1}{\vartheta} - 1\right),$$

we get

$$\begin{aligned}\left\| \frac{g-\mathbb{E}g}{\|\delta g\|_{\ell^2(\mathbb{N})}} \right\|_{\Phi_\varrho} &\le \mathcal{O}(1) \sup_{q>2} q^{1-1/\varrho} \zeta(1+\epsilon/(q-1))^{(q-1)/q}\, \mathbb{E}(\ell_0^{q+\epsilon})^{1/q} \\ &\quad + \mathcal{O}(1)\|\psi\|_{\ell^1(\mathbb{N})} \\ &\le \mathcal{O}(1) \sup_{q>2} \zeta \left(1+\frac{q(\vartheta-\varrho-\vartheta\varrho)}{\varrho(q-1)}\right)^{(q-1)/q} \|\ell_0\|_{\Phi_\vartheta}^{\vartheta(1-\varrho)/\varrho} \\ &\quad + \mathcal{O}(1)\|\psi\|_{\ell^1(\mathbb{N})} \\ &\le \mathcal{O}(1) \sup_{q>2} \zeta \left(\frac{\vartheta(1-\varrho)}{\varrho}\right)^{(q-1)/q} \|\ell_0\|_{\Phi_\vartheta}^{\vartheta(1-\varrho)/\varrho} + \mathcal{O}(1)\|\psi\|_{\ell^1(\mathbb{N})}\end{aligned}$$

since the function ζ is decreasing. Thus (18) is proved. The proof of the theorem is now complete. □

It is easy to obtain from Theorem 4 the following concentration inequalities.

Proposition 2 *Let n be an arbitrary positive integer.*

- *If $\mathbb{E}(\ell_0^{2p+\epsilon}) < \infty$ (for some $\epsilon > 0$), and $\|\psi\|_{\ell^1(\mathbb{N})} < \infty$, we have*

$$\mathbb{P}\{|g-\mathbb{E}g| > t\} \le C_p \frac{\|\delta g\|_{\ell^2(\mathbb{N})}^{2p}}{t^{2p}}, \tag{19}$$

  where $C_p \in]0,\infty[$, $p \in \mathbb{N}$, for any $g : A^n \to \mathbb{R}$.
- *Let $0 < \varrho < 1$. If $\|\psi\|_{\ell^1(\mathbb{N})} < \infty$, and $\|\ell_0\|_{\Phi_\vartheta} < \infty$ for some $\vartheta > \varrho/(1-\varrho)$, there exists a constant $c_{\varrho,\vartheta} \in]0,\infty[$ such that*

$$\mathbb{P}\{|g-\mathbb{E}g| > t\} \le 4\exp\left(-c_{\varrho,\vartheta}\, \frac{t^\varrho}{\|\delta g\|_{\ell^2(\mathbb{N})}^\varrho}\right), \tag{20}$$

  for any $g : A^n \to \mathbb{R}$.

Proof The proof of (19) is an immediate consequence of (17) applied to g and $-g$ and Chebychev's inequality.

For the proof of (20), we have for any $\lambda > 0$ using Chebychev's inequality

$$\begin{aligned}\mathbb{P}(g-\mathbb{E}g > t) &= \mathbb{P}\left(\frac{g-\mathbb{E}g}{\lambda\|\delta g\|_{\ell^2(\mathbb{N})}} > \frac{t}{\lambda\|\delta g\|_{\ell^2(\mathbb{N})}}\right) \\ &\le \mathbb{P}\left(\Phi_\varrho\left(\frac{g-\mathbb{E}g}{\lambda\|\delta g\|_{\ell^2(\mathbb{N})}}\right) > \Phi_\varrho\left(\frac{t}{\lambda\|\delta g\|_{\ell^2(\mathbb{N})}}\right)\right) \\ &\le \frac{1}{\Phi_\varrho\big(t/(\lambda\|\delta g\|_{\ell^2(\mathbb{N})})\big)}\, \mathbb{E}\left[\Phi_\varrho\left(\frac{g-\mathbb{E}g}{\lambda\|\delta g\|_{\ell^2(\mathbb{N})}}\right)\right].\end{aligned}$$

We now take $\lambda = \|(g - \mathbb{E}g)/\|\delta g\|_{\ell^2(\mathbb{N})}\|_{\Phi_\varrho}$. By definition, $\mathbb{E}\left[\Phi_\varrho\left(\frac{g-\mathbb{E}g}{\lambda\|\delta g\|_{\ell^2(\mathbb{N})}}\right)\right] = 1$. Thus we have

$$\mathbb{P}(g - \mathbb{E}g > t) \leq \frac{1}{\Phi_\varrho(t/\|g - \mathbb{E}g\|_{\Phi_\varrho})}.$$

Of course, the same inequality holds with $-g$. Applying (18) yields (20). The proposition is proved. □

In concrete applications of inequality (19) we have to check that $C_p < \infty$, otherwise the inequality is useless. To apply (20), we have to check that $c_\varrho > 0$. We will give an example of application below.

Inequality (19) is a "polynomial" concentration inequality whereas inequality (20) is a "stretched-exponential" concentration inequality.

Remark 5 The 4 in the r.h.s. of (20) is not optimal. It can be replaced by $2/(1-\epsilon)$ for any $\epsilon \in]0, 1[$.

2.4 Random fields

We now present the extension of our previous results to random fields. This requires mainly notational changes. We work with lattice spin systems. The configuration space is $\Omega = \{-, +\}^{\mathbb{Z}^d}$, endowed with the product topology. We could of course take any finite set A instead of $\{-, +\}$. For $\Lambda \subset \mathbb{Z}^d$ and $\sigma, \eta \in \Omega$ we denote $\sigma_\Lambda \eta_{\Lambda^c}$ the configuration coinciding with σ (resp. η) on Λ (resp. Λ^c). A local function $g : \Omega \to \mathbb{R}$ is such that there exists a finite subset $\Lambda \subset \mathbb{Z}^d$ such that for all σ, η, ω, $g(\sigma_\Lambda \omega_{\Lambda^c}) = g(\sigma_\Lambda \eta_{\Lambda^c})$.

For $\sigma \in \Omega$ and $x \in \mathbb{Z}^d$, σ^x denotes the configuration obtained from σ by "flipping" the spin at x. We denote $\delta_x g = \sup_\sigma |g(\sigma^x) - g(\sigma)|$ the variation of g at x. δg denotes the map $\mathbb{Z}^d \to \mathbb{R} : x \mapsto \delta_x g$.

We introduce the spiraling enumeration $\Gamma : \mathbb{Z}^d \to \mathbb{N}$ illustrated in the figure for the case $d = 2$.

We will use the abbreviation $(\leq x) = \{y \in \mathbb{Z}^d : \Gamma(y) \leq \Gamma(x)\}$ and similarly we introduce the abbreviations $(< x)$. By definition $\mathcal{F}_{\leq x}$ denotes the sigma-field generated by $\sigma(y)$, $y \leq x$ and $\mathcal{F}_{<0}$ denotes the trivial sigma-field.

For any local function $g : \Omega \to \mathbb{R}$, we have the analog decomposition as in (2):

$$g - \mathbb{E}g = \sum_{x \in \mathbb{Z}^d} V_x, \tag{21}$$

where

$$V_x := \mathbb{E}\left[g|\mathcal{F}_{\leq x}\right] - \mathbb{E}[g|\mathcal{F}_{<x}].$$

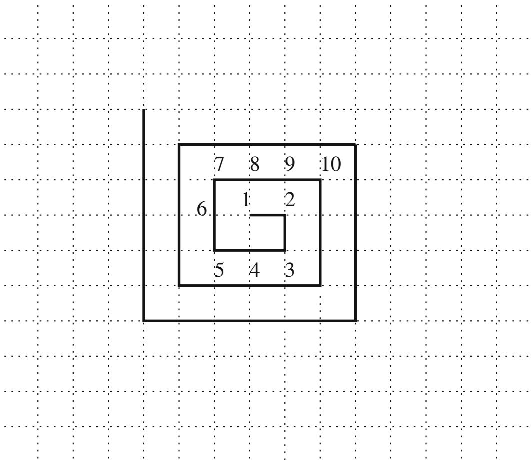

Fig. 1 "Spiral" ordering of $\mathbb{Z}^2$

The analog of the coupling matrix is the following matrix indexed by lattice sites $x, y \in \mathbb{Z}^d$

$$D^{\sigma}_{x,y} := \hat{\mathbb{P}}^{\sigma}_{x,+,-}\,\{X_1(y) \neq X_2(y)\}, \tag{22}$$

where $\hat{\mathbb{P}}^{\sigma}_{x,+,-}$ denotes the maximal coupling between the conditional measures $\mathbb{P}(\cdot|\sigma_{<x,+_x})$ and $\mathbb{P}(\cdot|\sigma_{<x,-_x})$. The notation "$+_x$" (resp. "$-_x$") means that at coordinate x in the configuration we put a "$+$" (resp. a "$-$").

We first consider the case of uniform decay of D. In that case, the exponential concentration inequality of Theorem 1 holds with the norm of $\ell_2(\mathbb{Z}^d)$, i.e., $\|\delta g\|_2^2 = \sum_{x\in\mathbb{Z}^d}(\delta_x g)^2$ (which is trivially finite since g is a local function).

Theorem 5 *Assume that*

$$\overline{D}_{x,y} := \sup_{\sigma} D^{\sigma}_{x,y} \tag{23}$$

is a bounded operator in $\ell_2(\mathbb{Z}^d)$. Then for all local functions g we have the following inequality

$$\mathbb{P}\,\{|g - \mathbb{E}g| \geq t\} \leq 2\exp\left(-\frac{2t^2}{\|\overline{D}\|^2_{\ell^2(\mathbb{Z}^d)}\;\|\delta g\|^2_{\ell^2(\mathbb{Z}^d)}}\right) \tag{24}$$

for all $t > 0$.

In the non-uniform case, Theorems 3, 4 and Proposition 2 extend immediately as follows. The analog of (12) is

$$D^{\sigma}_{x,y} \leq 1\!\!1\{\,\ell_x(\sigma) \geq |y - x|\} + \psi(|y - x|). \tag{25}$$

From now on, we assume that the distribution of ℓ_x is independent of x. We extend the matrix $\mathcal{D}$ defined in (11) by putting

$$\mathcal{D}^{(p)}_{x,y} := \mathbb{E}[(D^{\sigma}_{x,y})^p]^{1/p}$$

for $x, y \in \mathbb{Z}^d$.

Theorem 6 *For any local function g and for any $p \in \mathbb{N}$, we have*

$$\mathbb{E}\left[(g - \mathbb{E}g)^{2p}\right] \leq (20p)^p \; \|\mathcal{D}^{(2p)}\|^{2p}_{\ell^2(\mathbb{Z}^d)} \|\delta g\|^{2p}_{\ell^2(\mathbb{Z}^d)} .$$

Theorem 7 *For any local function g, for any $p \in \mathbb{N}$ and any $\epsilon > 0$, we have*

$$\begin{aligned}&\mathbb{E}\big[(g - \mathbb{E}g)^{2p}\big] \\ &\quad \leq (20p)^{2p} \left(\zeta(1+\epsilon/(2p-1))^{(2p-1)/2p} \mathbb{E}\big(\ell_0^{2pd+\epsilon}\big)^{1/2p} + \|\psi\|_{\ell^1(\mathbb{N})}\right)^{2p} \; \|\delta g\|^{2p}_{\ell^2(\mathbb{Z}^d)}\end{aligned} \tag{26}$$

where, ζ denotes Riemann's zeta function. For any $\vartheta > 0$, there is a constant $C_\vartheta > 0$, such that for any $\varrho < \vartheta/(1+\vartheta)$ satisfying $\zeta(\vartheta(1-\varrho)/\varrho) \geq 1$, we have

$$\left\| \frac{g - \mathbb{E}g}{\|\delta g\|_{\ell^2(\mathbb{Z}^d)}} \right\|_{\Phi_\varrho} \leq C_\vartheta \left(\zeta\left(\frac{\vartheta(1-\varrho)}{\varrho}\right) \|\ell_0^d\|^{\vartheta(1-\varrho)/\varrho}_{\Phi_\vartheta} + \|\psi\|_{\ell^1(\mathbb{N})} \right). \tag{27}$$

Proposition 3 *For any local function g we have the inequalities:*

- *If $\mathbb{E}(\ell_0^{2pd+\epsilon}) < \infty$ (for some $\epsilon > 0$), and $\|\psi\|_{\ell^1(\mathbb{N})} < \infty$, we have*

  $$\mathbb{P}\{|g - \mathbb{E}g| > t\} \leq C_p \frac{\|\delta g\|^{2p}_{\ell^2(\mathbb{Z}^d)}}{t^{2p}}$$

  where $C_p \in]0, \infty[$, $p \in \mathbb{N}$.
- *Let $0 < \varrho < 1$. If $\|\psi\|_{\ell^1(\mathbb{N})} < \infty$, and $\|\ell_0^d\|_{\Phi_\vartheta} < \infty$ for some $\vartheta > \varrho/(1-\varrho)$, there exists a constant $c_{\varrho,\vartheta} \in]0, \infty[$ such that*

  $$\mathbb{P}\{|g - \mathbb{E}g| > t\} \leq 4 \exp\left(-c_{\varrho,\vartheta} \; \frac{t^\varrho}{\|\delta g\|^\varrho_{\ell^2(\mathbb{Z}^d)}} \right).$$

Remark 6 It is immediate to extend the previous inequalities to integrable functions g belonging to the closure of the set of local functions with the norm $|||g||| := \|\delta g\|_{\ell^2(\mathbb{Z}^d)}$.

2.5 Existence of a coupling by bounding the variation

We continue with random fields and state a proposition which says that if we have an estimate of the form

$$V_x \leq (D\delta g)_x$$

for some matrix D, then there exists a coupling with coupling matrix $\hat{D}$ such that its matrix elements decay at least as fast as the matrix elements of D. We formulate the proposition more abstractly:

Proposition 4 *Suppose that $\mathbb{P}$ and $\mathbb{Q}$ are probability measures on Ω and for all $g : \Omega \to \mathbb{R}$ we have the estimate*

$$\left|\mathbb{E}_{\mathbb{P}}[g] - \mathbb{E}_{\mathbb{Q}}[g]\right| \leq \sum_{x\in\mathbb{Z}^d} \rho(x)\delta_x g \tag{28}$$

for some "weights" $\rho : \mathbb{Z}^d \to \mathbb{R}^+$. Suppose $\varphi : \mathbb{Z}^d \to \mathbb{R}^+$ is such that

$$\sum_{x\in\mathbb{Z}^d} \rho(x)\varphi(x) < \infty\,.$$

Then there exists a coupling $\hat{\mu}$ of $\mathbb{P}$ and $\mathbb{Q}$ such that

$$\sum_{x\in\mathbb{Z}^d} \hat{\mu}\,\{X_1(x) \neq X_2(x)\}\,\varphi(x) \leq \sum_{x\in\mathbb{Z}^d} \varphi(x)\rho(x) < \infty.$$

Proof Let $B_n := [-n,n]^d \cap \mathbb{Z}^d$. Define the "cost" function

$$C_n^{\varphi}(\sigma,\sigma') := \sum_{x\in B_n} |\sigma_x - \sigma'_x|\,\varphi(x).$$

Denote by $\mathbb{P}_n$, respectively, $\mathbb{Q}_n$, the joint distribution of $\{\sigma_x, x \in B_n\}$ under $\mathbb{P}$, respectively, $\mathbb{Q}$. Consider the class of functions

$$\mathcal{G}_{C_n^{\varphi}} := \{g|\; g \in \mathcal{F}_{B_n},\, |g(\sigma) - g(\sigma')| \leq \sum_{x\in\mathbb{Z}^d} \varphi(x)1\!\!1\{\sigma_x \neq \sigma'_x\},\, \forall\sigma,\sigma' \in \Omega\}\,.$$

It is obvious from the definition that $g \in \mathcal{G}_{C_n^{\varphi}}$, if, and only if, g is $\mathcal{F}_{B_n}$-measurable and

$$(\delta_x g)(\sigma) \leq \varphi(x) \quad \forall x \in B_n, \quad \forall\sigma \in \Omega\,.$$

Therefore, if (28) holds, then for all $g \in \mathcal{G}_{C_n^{\varphi}}$,

$$\left|\mathbb{E}_{\mathbb{P}}[g] - \mathbb{E}_{\mathbb{Q}}[g]\right| \leq \sum_{x\in\mathbb{Z}^d} \rho(x)\delta_x g \leq \sum_{x\in\mathbb{Z}^d} \rho(x)\varphi(x)\,.$$

Hence, by the Kantorovich–Rubinstein duality theorem [18], there exists a coupling $\hat{\mu}_n$ of $\mathbb{P}_n$ and $\mathbb{Q}_n$ such that

$$\mathbb{E}_{\hat{\mu}_n}\left(C_n^{\varphi}(\sigma,\sigma')\right) = \mathbb{E}_{\hat{\mu}_n}\left(\sum_{x\in B_n}\varphi(x)\mathbb{1}\{X_1(x)\neq X_2(x)\}\right) \leq \sum_{x\in\mathbb{Z}^d}\varphi(x)\rho(x).$$

By compactness (in the weak topology), there exists a subsequence along which $\hat{\mu}_n$ converges weakly to some probability measure $\hat{\mu}$. For any $k \leq n$, we have

$$\begin{aligned}&\mathbb{E}_{\hat{\mu}_n}\left(\sum_{x\in B_k}\varphi(x)\mathbb{1}\{X_1(x)\neq X_2(x)\}\right)\\ &\quad\leq \mathbb{E}_{\hat{\mu}_n}\left(\sum_{x\in B_n}\varphi(x)\mathbb{1}\{X_1(x)\neq X_2(x)\}\right) \leq \sum_{x\in\mathbb{Z}^d}\varphi(x)\rho(x).\end{aligned}$$

Therefore, taking the limit $n\to\infty$ along the above subsequence yields

$$\mathbb{E}_{\hat{\mu}}\left(\sum_{x\in B_k}\varphi(x)\mathbb{1}\{X_1(x)\neq X_2(x)\}\right) \leq \sum_{x\in\mathbb{Z}^d}\varphi(x)\rho(x)\,.$$

We now take the limit $k\to\infty$ and use monotonicity to conclude that

$$\mathbb{E}_{\hat{\mu}}\left(\sum_{x\in\mathbb{Z}^d}\varphi(x)\mathbb{1}\{X_1(x)\neq X_2(x)\}\right) \leq \sum_{x\in\mathbb{Z}^d}\varphi(x)\rho(x).$$

□

We shall illustrate below this proposition with the example of Gibbs random fields at high-temperature under the Dobrushin uniqueness condition.

3 Examples

3.1 High-temperature Gibbs measures

For the sake of convenience, we briefly recall a few facts about Gibbs measures. We refer to [9] for details.

A finite-range potential (with range R) is a family of functions $U(A,\sigma)$ indexed by finite subsets A of $\mathbb{Z}^d$ such that the value of $U(A,\sigma)$ depends only on σ_A and such that $U(A,\sigma)=0$ if $\mathrm{diam}(A)>R$. If $R=1$ then the potential is nearest-neighbor.

The associated finite-volume Hamiltonian with boundary condition η is then given by

$$H_\Lambda^\eta(\sigma) = \sum_{A\cap\Lambda\neq\emptyset} U(A, \sigma_\Lambda \eta_{\Lambda^c}) \,.$$

The specification is then defined as

$$\gamma_\Lambda(\sigma|\eta) = \frac{\mathrm{e}^{-H_\Lambda^\eta(\sigma)}}{Z_\Lambda^\eta}.$$

We then say that $\mathbb{P}$ is Gibbs measure with potential U if $\gamma_\Lambda(\sigma|\cdot)$ is a version of the conditional probability $\mathbb{P}(\sigma_\Lambda|\mathcal{F}_{\Lambda^c})$.

Before we state our result, we need some notions from [8]. What we mean by "high temperature" will be an estimate on the variation of single-site conditional probabilities, which will imply a uniform estimate for disagreement percolation. For $y \in \mathbb{Z}^d$, let

$$p_y := 2 \sup_{\sigma,\sigma'} \left| \mathbb{P}(\sigma_y = +|\sigma_{\mathbb{Z}^d\setminus y}) - \mathbb{P}(\sigma'_y = +|\sigma'_{\mathbb{Z}^d\setminus y}) \right| \,.$$

Writing $\mathbf{p}$ for $(p_y)_y$, let $\nu_\mathbf{p}$ denote the Bernoulli measure on $\{-,+\}^{\mathbb{Z}^d}$ with $\nu_\mathbf{p}(\{X(y) = +\}) = p_y$, and ν_{p_y} its single-site marginal.

From [8, Theorem 7.1] it follows that there exists a coupling $\mathbb{P}^\sigma_{x,+,-}$ of the conditional distributions $\mathbb{P}(\cdot|\sigma_{<x}, +_x)$ and $\mathbb{P}(\cdot|\sigma_{<x}, -_x)$ such that under this coupling

1. For $y > x$, the event $X_1(y) \neq X_2(y)$ coincides with the event that there exists a path $\gamma \subset \mathbb{Z}^d \setminus (< x)$ from x to y such that, for all $z \in \gamma$, $X_1(z) \neq X_2(z)$. We denote this event by "$x \not\leftrightarrow y$".
2. The distribution of $\mathbb{1}\{X_1(y) \neq X_2(y)\}$ for $y \in \mathbb{Z}^d \setminus (\leq x)$ under $\mathbb{P}^\sigma_{x,+,-}$ is dominated by the product measure

$$\prod_{y\in\mathbb{Z}^d\setminus(\leq x)} \nu_{p_y} \,.$$

Let $p_c = p_c(d)$ be the critical percolation threshold for site-percolation on $\mathbb{Z}^d$. It then follows from statements 1 and 2 above that, if

$$\sup\{p_y : y \in \mathbb{Z}^d\} < p_c \tag{29}$$

then we have the uniform estimate

$$\mathbb{P}^{\sigma}_{x,+,-}\{X_1(y) \neq X_2(y)\} \leq \prod_{y \in \mathbb{Z}^d \setminus (\leq x)} \nu_{p_y}(x \leftrightarrow y) \leq \mathrm{e}^{-c|x-y|}. \tag{30}$$

Then we can apply Theorem 5 to obtain

Theorem 8 *Let U be a nearest-neighbor potential such that* (29) *holds. Then for the coupling matrix* (22) *we have the uniform estimate*

$$D^{\sigma}_{x,y} \leq e^{-C|x-y|}$$

for some $C > 0$*. Hence we have the following exponential concentration inequality: for any local function g and for all* $t > 0$

$$\mathbb{P}\{|g - \mathbb{E}g| \geq t\} \leq 2 \exp\left(-\frac{2t^2}{\frac{1}{1-e^{-2C}}\ \|\delta g\|^2_{\ell^2(\mathbb{Z}^d)}}\right).$$

Remark 7 Theorem 8 can easily be extended to any finite-range potential.

Theorem 8 was obtained in [11] in the Dobrushin's uniqueness regime [9, Chapter 8] using a different approach. The high-temperature condition which we use here is sometimes less restrictive than Dobrushin's uniqueness condition, but sometimes it is more restrictive. However, Dobrushin's uniqueness condition is not limited to finite-range potentials. We now apply Proposition 4 to show that in the Dobrushin's uniqueness regime, there does exist a coupling of $\mathbb{P}(\cdot|\sigma_{<x,+_x})$ and $\mathbb{P}(\cdot|\sigma_{<x,-_x})$ such that the elements of the associated coupling matrix decay at least as fast as the elements of the Dobrushin's matrix. The Dobrushin's uniqueness condition is based on the matrix

$$C_{x,y} := 2 \sup_{\sigma,\sigma' : \sigma_{\mathbb{Z}^d \setminus y} = \sigma'_{\mathbb{Z}^d \setminus y}} \left| \mathbb{P}(\sigma_x = + | \sigma_{\mathbb{Z}^d \setminus x}) - \mathbb{P}(\sigma_x = + | \sigma'_{\mathbb{Z}^d \setminus x}) \right|.$$

This condition is defined by requiring that

$$\sup_{x \in \mathbb{Z}^d} \sum_{y \in \mathbb{Z}^d} C_{x,y} < 1$$

and the Dobrushin matrix is then defined as

$$\Delta_{x,y} := \sum_{n \geq 0} C^n_{x,y}.$$

We now have the following proposition:

Proposition 5 *Assume that the Dobrushin uniqueness condition holds. For any $\varphi : \mathbb{Z}^d \to \mathbb{R}^+$ such that for any $x \in \mathbb{Z}^d$,*

$$\sum_{y\in\mathbb{Z}^d} \varphi(y)\Delta_{y,x} < \infty.$$

Then there exists a coupling $\hat{\mathbb{P}}^{\sigma}_{x,+,-}$ of $\mathbb{P}(\cdot|\sigma_{<x,+_x})$ and $\mathbb{P}(\cdot|\sigma_{<x,-_x})$ such that

$$\sum_{y\in\mathbb{Z}^d} \varphi(y)\, \hat{\mathbb{P}}^{\sigma_{<x}}_{x,+,-}\{X_1(y) \neq X_2(y)\} < \infty\,.$$

Proof From [11, Lemma 1], we have the estimate

$$\left|\int g(\eta)\,\mathbb{P}(d\eta|\sigma_{<x,+_x}) - \int g(\eta)\,\mathbb{P}(d\eta|\sigma_{<x,-_x})\right| \le \sum_{y\in\mathbb{Z}^d}(\mathbb{1}_{x,y} + \Delta_{y,x})\delta_y g$$

(where $\mathbb{1}_{x,y}$ denotes the Kronecker symbol).

We can apply Proposition 4 to conclude the proof. □

As an example we mention that if the potential is finite-range and translation-invariant and satisfies the Dobrushin uniqueness condition, we have for large enough $|x-y|$

$$\Delta_{y,x} \le \mathrm{e}^{-c|x-y|}$$

and hence there exists a coupling $\hat{\mathbb{P}}^{\sigma_{<x}}_{x,+,-}$ such that

$$\hat{\mathbb{P}}^{\sigma_{<x}}_{x,+,-}\{X_1(y) \neq X_2(y)\} \le \mathrm{e}^{-c'|x-y|}$$

for all $c' < c$ and large enough $|x-y|$.

Unfortunately, we are not able to construct explicitly such a coupling.

3.2 The low-temperature Ising model

It is clear that for the Ising model in the phase coexistence region, no exponential concentration inequalities can hold. Indeed, this would contradict the surface-order large deviation bounds for the magnetization in that regime (see e.g. [10] and references therein). Nevertheless, we shall show that we can control all moments and obtain stretched-exponential inequalities (which are compatible with large deviation bounds).

We consider the low-temperature plus phase of the Ising model on $\mathbb{Z}^d$, $d \ge 2$. This is a probability measure $\mathbb{P}^+_\beta$ on lattice spin configurations $\sigma \in \Omega$, defined as the weak limit as $\Lambda \uparrow \mathbb{Z}^d$ of the following finite-volume measures:

$$\mathbb{P}^{+}_{\Lambda,\beta}(\sigma_\Lambda) = \exp\left(\beta \sum_{<xy>\in\Lambda} \sigma_x\sigma_y + \beta \sum_{<xy>,x\in\partial\Lambda,\ y\notin\Lambda} \sigma_x\right) \Big/ Z^{+}_{\Lambda,\beta}, \tag{31}$$

where $\beta \in \mathbb{R}^+$ is the inverse temperature, and $Z^{+}_{\Lambda,\beta}$ is the partition function. In (31) $\langle xy\rangle$ denotes nearest neighbor bonds and $\partial\Lambda$ the inner boundary, i.e., the set of those $x \in \Lambda$ having at least one neighbor $y \notin \Lambda$. The existence of the limit $\Lambda \uparrow \mathbb{Z}^d$ of $\mathbb{P}^{+}_{\Lambda,\beta}$ is by a standard and well-known monotonicity argument, see e.g. [9].

For any $\eta \in \Omega$, $\Lambda \subset \mathbb{Z}^d$ we denote by $\mathbb{P}^{\eta}_{\Lambda,\beta}$ the corresponding finite-volume measure with boundary condition η:

$$\mathbb{P}^{\eta}_{\Lambda,\beta}(\sigma_\Lambda) = \exp\left(\beta \sum_{<xy>\in\Lambda} \sigma_x\sigma_y + \beta \sum_{x\in\Lambda,\ y\notin\Lambda} \sigma_x\eta_x\right) \Big/ Z^{\eta}_{\Lambda,\beta}\,.$$

Later on we will have to choose β large enough, in particular, greater than the critical inverse temperature β_c ($\beta < \beta_c$ implies uniqueness of the infinite-volume measure).

We can now formulate our results on arbitrary local functions for the low-temperature Ising model.

Theorem 9 *Let $\mathbb{P} = \mathbb{P}^{+}_{\beta}$ be the plus phase of the low-temperature Ising model defined above. There exists $\beta_0 > \beta_c$, such that for all $\beta > \beta_0$, for any local function g, we have the following inequalities:*

- *For all $p \in \mathbb{N}$, there exists a constant $C_p \in]0,\infty[$ such that*

$$\mathbb{E}\left[(g-\mathbb{E}g)^{2p}\right] \le C_p\ \|\delta g\|^{2p}_{\ell^2(\mathbb{Z}^d)}\,.$$

Consequently, for all $t > 0$, we have the concentration inequalities

$$\mathbb{P}\{|g-\mathbb{E}g| > t\} \le C_p\ \frac{\|\delta g\|^{2p}_{\ell^2(\mathbb{Z}^d)}}{t^{2p}}\cdot$$

- *Moreover, there exists $0 < \varrho(\beta) < 1$, such that for any $0 < \varrho < \varrho(\beta)$ there is a constant $K_\varrho > 0$, such that we have, for any local function g,*

$$\|g-\mathbb{E}g\|_{\Phi_\varrho} \le K_\varrho\ \|\delta g\|_{\ell^2(\mathbb{Z}^d)}\,.$$

Consequently, there exists a constant $c_\varrho \in]0,\infty[$ such that, for all $t > 0$,

$$\mathbb{P}\{|g-\mathbb{E}g| > t\} \le 4\exp\left(-c_\varrho\ \frac{t^\varrho}{\|\delta g\|^{\varrho}_{\ell^2(\mathbb{Z}^d)}}\right)\cdot$$

Proof This theorem is an application of Theorem 7 and Proposition 3. All we have to do is to obtain the bound (25) with good decay properties for the tail of the distribution of ℓ_0 to ensure the finiteness of $\mathbb{E}(\ell_0^{2pd+\epsilon})$, $\|\ell_0^d\|_{\Phi_\vartheta}$, and of $\|\psi\|_{\ell^1(\mathbb{N})}$. This is an immediate consequence of the next proposition. □

Proposition 6 *Let $\mathbb{P} = \mathbb{P}_\beta^+$ be the plus phase of the low-temperature Ising model. There exists $\beta_0 > \beta_c$ such that for all $\beta > \beta_0$, the inequality* (25) *holds together with the estimate*

$$\psi(n) \le C\mathrm{e}^{-\mathrm{cn}}$$

for all $n \in \mathbb{N}$ and

$$\mathbb{P}\{\ell_0 \ge n\} \le C'\mathrm{e}^{-c'n^\alpha}$$

for some $c, c', C, C' > 0$ and $0 < \alpha \le 1$.

Proof We shall make a coupling of the conditional measures $\mathbb{P}(\cdot|\sigma_{<x,+_x})$ and $\mathbb{P}(\cdot|\sigma_{<x,-_x})$. This coupling already appeared in [23] (see also [8]). Both conditional measures are a distribution of a random field ω_y, $y \notin (\le x)$. We start with the first site $y_1 > x$ according to the order induced by Γ (see Sect. 2.4). We generate $X_1(y_1)$ and $X_2(y_1)$ as a realization of the maximal coupling between $\mathbb{P}(\sigma_{y_1} = \cdot|\sigma_{<x,+_x})$ and $\mathbb{P}(\sigma_{y_1} = \cdot|\sigma_{<x,-_x})$. Given that we have generated $X_1(y), X_2(y), \ldots, X_1(y_n), X_2(y_n)$ for $y = y_1, \ldots, y_n$, we generate $X_1(y_{n+1})$, $X_2(y_{n+1})$ for the smallest $y_{n+1} > y_n$ as a realization of the maximal coupling between

$$\mathbb{P}(\sigma_{y_{n+1}} = \cdot|X_1(y_1)\cdots X_1(y_n)\sigma_{<x,+_x}) \quad \text{and} \quad \mathbb{P}(\sigma_{\mathrm{y_{n+1}}} = \cdot|\mathrm{X_2(y_1)\cdots X_2(y_n)}\sigma_{<\mathrm{x},-_\mathrm{x}}).$$

By the Markov property of $\mathbb{P}$ we have the following: if there exists a contour separating y from x such that for all sites z belonging to that contour we have $X_1(z) = X_2(z)$, then $X_1(y) = X_2(y)$. The complement of this event (of having such a contour) is contained in the event that there exists a path of disagreement from x to y, i.e., a path $\gamma \subset \mathbb{Z}^d \setminus (< x)$ such that for all $z \in \gamma$, $X_1(z) \ne X_2(z)$. Denote that event by E_{xy}. Clearly its probability is bounded from above by the probability of the same event in the product coupling. In turn the event E_{xy} is contained in the event E_{xy}^+ that there exists a path γ from x to y in $\mathbb{Z}^d \setminus (< x)$ such that for all $z \in \gamma$, $(X_1(z), X_2(z)) \ne (+,+)$. In [14] (more precisely p. 531 estimate (48) and p. 535 estimate (63)), the probability of that event in the product coupling is precisely estimated from above by

$$Ce^{-c|x-y|} + 1\!\!1\{\ell_x(\sigma) \ge |x-y|\} \tag{32}$$

for some $C, c > 0$, where $\ell_x(\sigma)$ is an unbounded function of σ with tail estimate

$$\mathbb{P}(\ell_x(\sigma) \ge n) = \mathbb{P}(\ell_0(\sigma) \ge n) \le C'e^{-c'n^\alpha}$$

for some $C', c' > 0$ and $0 < \alpha < 1$. For the reader's convenience, we briefly comment on these estimates. The idea is that the conditional measure $\mathbb{P}(\cdot|\xi_{\leq x})$ resembles the original unconditioned plus phase (in $\mathbb{Z}^d \setminus (\leq x)$) provided ξ contains "enough" pluses. "Containing enough pluses" is exactly quantified by the random variable $\ell_x(\xi)$: $(\ell_x(\xi) \leq n)$ is the event that for all self-avoiding path γ of length at least n, the magnetization along γ,

$$m_\gamma(\xi) := \frac{1}{|\gamma|} \sum_{z \in \gamma} \xi_z$$

is close "enough to one". If this is the case then under the conditional measure we still have a Peierls' estimate, which produces the exponential term in (32). We refer to [14] for more details. □

Acknowledgments We wish to thank a referee for a careful reading of our manuscript, as well as for the impetus he gave us to include stretched-exponential concentration inequalities which we did not consider in the primitive version of this work.

References

1. Bobkov, S.G., Götze, F.D.: Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J. Funct. Anal. **163**(1), 1–28 (1999)
2. Chazottes, J.-R., Collet, P., Schmitt, B.: Devroye inequality for a class of non-uniformly hyperbolic dynamical systems. Nonlinearity **18**(5), 2341–2364 (2005)
3. Collet, P., Martínez, S., Schmitt, B.: Exponential inequalities for dynamical measures of expanding maps of the interval. Probab. Theor. Relat. Fields **123**, 301–322 (2002)
4. Devroye, L.: Exponential inequalities in nonparametric estimation. Nato ASI series C, Math. Phys. Sci. vol. 335, pp. 31–44. Kluwer, Dordrecht (1991)
5. Devroye, L., Lugosi, G.: Combinatorial methods in density estimation. Springer Series in Statistics. Springer, Berlin Heidelberg New York (2001)
6. Dudley, R.M.: Real analysis and probability. In: Cambridge studies in advanced mathematics (2002)
7. Garsia, A.M.: Martingale inequalities: Seminar notes on recent progress. In: Benjamin, W.A. (ed.)Mathematics Lecture Notes Series. Reading (1973)
8. Georgii, H.-O.: Gibbs Measures and Phase Transitions. Walter de Gruyter & Co., Berlin (1988)
9. Georgii, H.O., Häggström, O., Maes, C.: The Random Geometry of Equilibrium Phases. In: Domb, C., Lebowitz, J.L. (eds.) Phase transitions and critical phenomena, vol. 18, pp. 1–142. Academic, London (2001)
10. Ioffe, D.: Exact large deviation bounds up to T_c for the Ising model in two dimensions. Probab. Theory Relat. Fields **102**(3), 313–330 (1995)
11. Külske, C.: Concentration inequalities for functions of Gibbs fields with applications to diffraction and random Gibbs measures. Commun. Math. Phys. **239**, 29–51 (2003)
12. Ledoux, M.: The concentration of measure phenomenon. In: Mathematical Surveys and Monographs, vol. 89. American Mathematical Society, Providence R.I. (2001)
13. Lindvall, T.: Lectures on the coupling method. Wiley Series in Probab. and Math. Stat. Wiley, (1992)
14. Maes, C., Redig, F., Shlosman, S., Van Moffaert, A.: Percolation, Path large deviations and weak Gibbsianity. Commun. Math. Phys. **209**, 517–545 (2000)
15. Marton, K.: Measure concentration for a class of random processes. Probab. Theory Relat. Fields **110**(3), 427–439 (1998)
16. Marton, K.: Measure concentration and strong mixing. Stud. Sci. Math. Hungar. **40**(1–2) 95–113 (2003)

17. Marton, K.: Measure concentration for Euclidean distance in the case of dependent random variables. Ann. Probab. **32**(3B), 2526–2544 (2004)
18. Rachev, S.T.: Probability metrics and the stability of stochastic models. In: Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. Wiley Chicheter (1991)
19. Rao, M.M., Ren, Z.D.: Applications of Orlicz spaces. In: Monographs and Textbooks in Pure and Applied Mathematics, vol. 250. Marcel Dekker, Inc., New York (2002)
20. Rio, E.: Inégalités de Hoeffding pour les fonctions lipschitziennes de suites dépendantes. C. R. Acad. Sci. Paris Sér. I Math. **330**(10) 905–908 (2000)
21. Samson, P.-M.: Concentration of measure inequalities for Markov chains and Φ-mixing processes. Ann. Probab. **28**(1), 416–461 (2000)
22. Su, Z.: Central limit theorems for random processes with sample paths in exponential Orlicz spaces. Stochast. Process. Appl. **66**(1), 1–20 (1997)
23. van den Berg, J., Maes, C.: Disagreement percolation in the study of Markov fields. Ann. Probab. **22**, 749–763 (1994)
24. Yukich, J.E.: Probability theory of classical Euclidean optimization problems. In: Lecture Notes in Mathematics vol. 1675. Springer, Berlin Heidelberg New York, (1998)
25. Zygmund, A.: Trigonometric series, vol. I and II, 3rd edn. Cambridge Mathematical Library. Cambridge University Press, Cambridge (2002)